Ogden Air Logistics Center

Can We Trust Our Machines?

Daniel P. Stormont

Embedded Systems Engineer

520 SMXS/MXDEA

801-586-1374

daniel.stormont@hill.af.mil

Overview

- **Why Trust Matters**
- **Autonomy Defined**
- **Trust Defined**
- **Research Objective**
- **Identifying Trust Factors**
- **Simulating Trust**
- **Human-Robot Trust Game Prototype**
- **Web-based Game**
- **Identifying Human-Agent Trust KPPs**
- **Trust in Reverse?**

Why Trust Matters

- **Congressional Mandate**
  - **The National Defense Authorization Act for Fiscal Year 2001, Public Law 106-398, Congress mandated in Section 220 that "It shall be a goal of the Armed Forces to achieve the fielding of unmanned, remotely controlled technology such that… by 2015, one-third of the operational ground combat vehicles are unmanned."[1]**

Why Trust Matters

- **Secretary of Defense Gates' Comments**
  - **On 21 April 2008, Secretary Gates made the following comment about unmanned systems**

    "Unmanned systems cost much less and offer greater loiter times than their manned counterparts, making them ideal for many of today's tasks. Today, we now have more than 5,000 UAVs, a 25-fold increase since 2001. But in my view, we can do and we should do more to meet the needs of men and women fighting in the current conflicts while their outcome may still be in doubt. My concern is that our services are still not moving aggressively in wartime to provide resources needed now on the battlefield. I've been wrestling for months to get more intelligence, surveillance and reconnaissance assets into the theater. Because people were stuck in old ways of doing business, it's been like pulling teeth."[2]
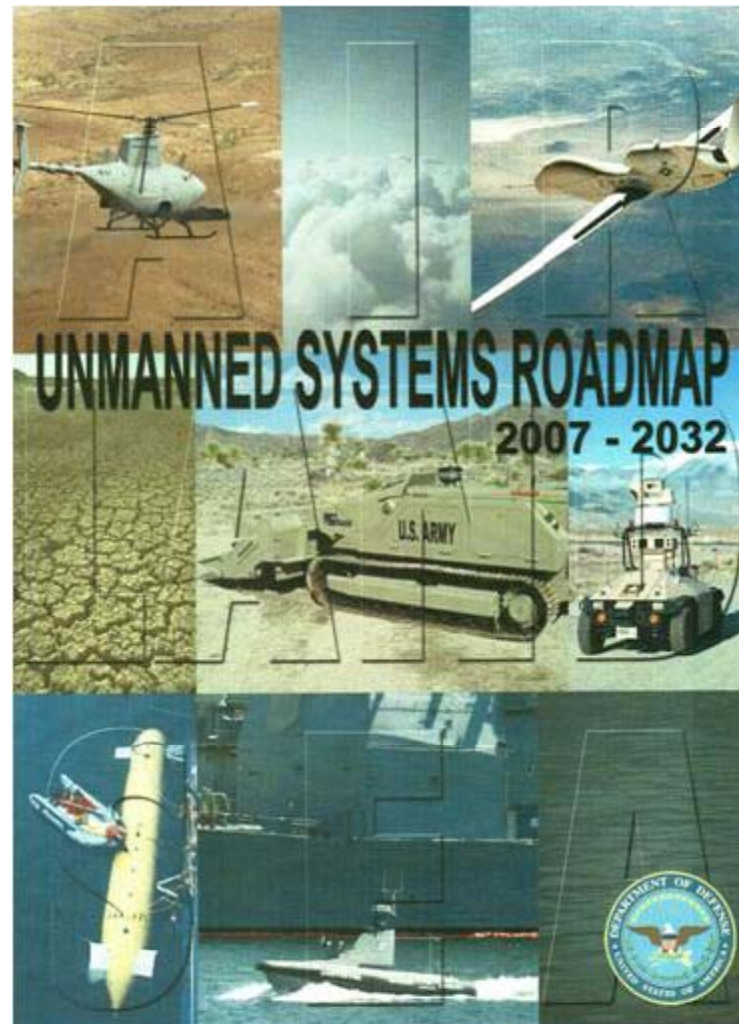
Why Trust Matters

UNMANNED SYSTEMS ROADMAP
2007 - 2032

U.S. ARMY

Why Trust Matters

- **P.W. Singer of the Brookings Institution stated the following in an article published this year**
  - *"So, despite what one article called 'all the lip service paid to keeping a human in the loop,' the cold, hard, metallic reality is that autonomous armed robots are coming to war. They simply make too much sense to the people that matter."* [4]

Autonomy Defined

- **The American Heritage Dictionary defines *autonomy* as**
  - *"Independence."* [5]
- **Wikipedia defines an *autonomous agent* as**
  - *"A system situated in, and part of, an environment, which senses that environment, and acts on it, over time, in pursuit of its own agenda. This agenda evolves from drives (or programmed goals). The agent acts to change the environment and influences what it senses at a later time."*
  - *"Non-biological examples include intelligent agents, autonomous robots, and various software agents, including artificial life agents, and many computer viruses."* [6]

Autonomy Defined

- **There are varying degrees of autonomy**
  - **None – System is completely manually controlled**
  - **Partial – Some functions are automated**
  - **Sliding – The amount of autonomy is selectable**
  - **Full – The system operates entirely without human control**
- **With any autonomy the user gives up some level of control**
- **Wikipedia has a good definition of the abilities of a fully autonomous robot**
  - *"A fully autonomous robot has the ability to*
    - *Gain information about the environment*
    - *Work for an extended period without human intervention*
    - *Move either all or part of itself throughout its operating environment without human assistance*
    - *Avoid situations that are harmful to people, property, or itself unless those are part of its design specifications."*[7]

Autonomy Defined

■ **Examples of Degrees of Autonomy**

Increasing Autonomy

Full Autonomy (ALCM)

Boeing Photo

Sliding Autonomy (UAS)

US Air Force Photo

Partial Autonomy (ISS assembly robot)

NASA Photo

No Autonomy (Remote-control EOD robot)

US Air Force Photo

Trust Defined

- **The "classic" definition of trust comes from Diego Gambetta**
  - "trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both *before* he can monitor such action (or independently of his capacity ever to be able to monitor it) *and* in a context in which it affects *his own* action. When we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him."[8]

Research Objective

- **Assuming that autonomous agents are coming to hazardous environments, like disaster areas and combat zones, how can we trust them?**

- **This research project is attempting to identify the factors that contribute to trust in autonomous agents, in order to develop a set of key performance parameters (KPPs) for trusted agents**
  - **These KPPs can be used by intelligent agent developers to validate the trustworthiness of their agents and to convince users of their trustworthiness**
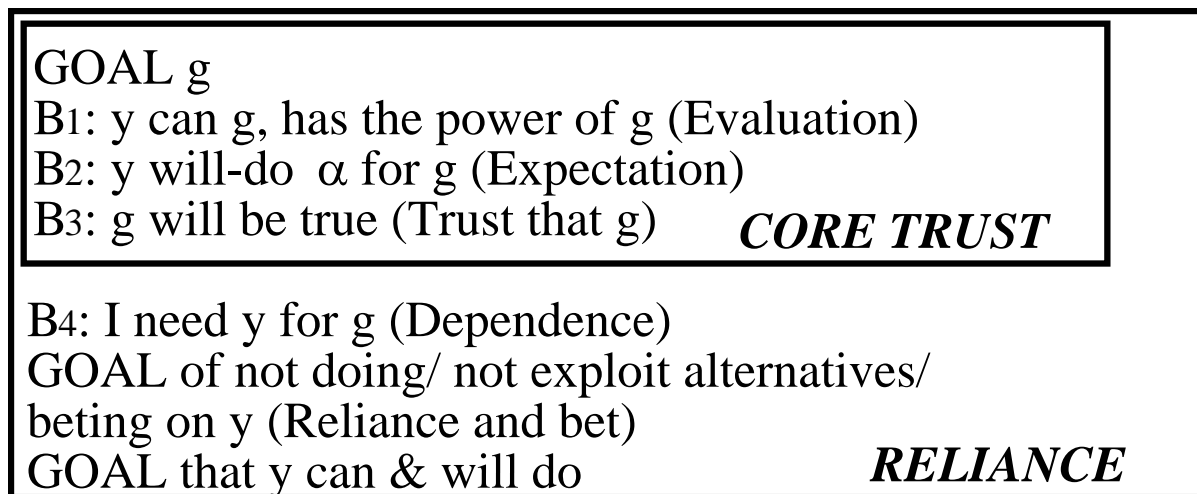
Research Objective

- **Three elements to the research**
  - **Identify factors of trust published in the literature**
  - **Simulate human-robot interactions**
  - **Collect data on human interactions with autonomous agents**
- **The ultimate objective is to develop a process that can be validated in field trials**

Identifying Trust Factors

- **Cristiano Castelfranchi and Rino Falcone argue that**

  Only agents endowed with goals and beliefs (cognitive agents) can "trust" another agent[9]

- **They define the elements of trust with the following diagram[9]**

GOAL g
B1: y can g, has the power of g (Evaluation)
B2: y will-do α for g (Expectation)
B3: g will be true (Trust that g)   ***CORE TRUST***

B4: I need y for g (Dependence)
GOAL of not doing/ not exploit alternatives/
beting on y (Reliance and bet)
GOAL that y can & will do   ***RELIANCE***

Mental ingredients of TRUST

Identifying Trust Factors

- **Sarvapali Ramchurn, et al., identified two principle components of trust in an agent[10]**
    - **Confidence – Do I believe the agent can perform the desired task?**
    - **Reputation – Has this agent been successful in the past and have others trusted this agent, with good results?**

Identifying Trust Factors

- **Karen Fullam and Suzanne Barber focused on the importance of reputation when dealing with agents (either human or artificial) in the development of the ART Testbed[11]**

Identifying Trust Factors

- **Just in these three research papers, the following potential trust factors were identified**
  - **Evaluation**
  - **Expectation**
  - **Trust**
  - **Dependence**
  - **Reliance**
  - **Bet**
  - **Confidence**
  - **Reputation**
- **Which factors are the key factors?**

Simulating Trust

- **Trust of autonomous systems is an active area of research**
- **Much of the research uses simulations to explore issues of trust and to compare approaches**
- **The simulations tend to take two forms**
  - **Simulations designed specifically for examining issues of trust**
  - **Simulations that were designed for other areas of research that have been extended or adapted to trust research**
- **The following are some examples of both types of simulation**

Simulating Trust

- **Agent Reputation and Trust (ART) Testbed[12]**
  - **A simulation of multiple art appraising agents specifically designed for agent trust research**
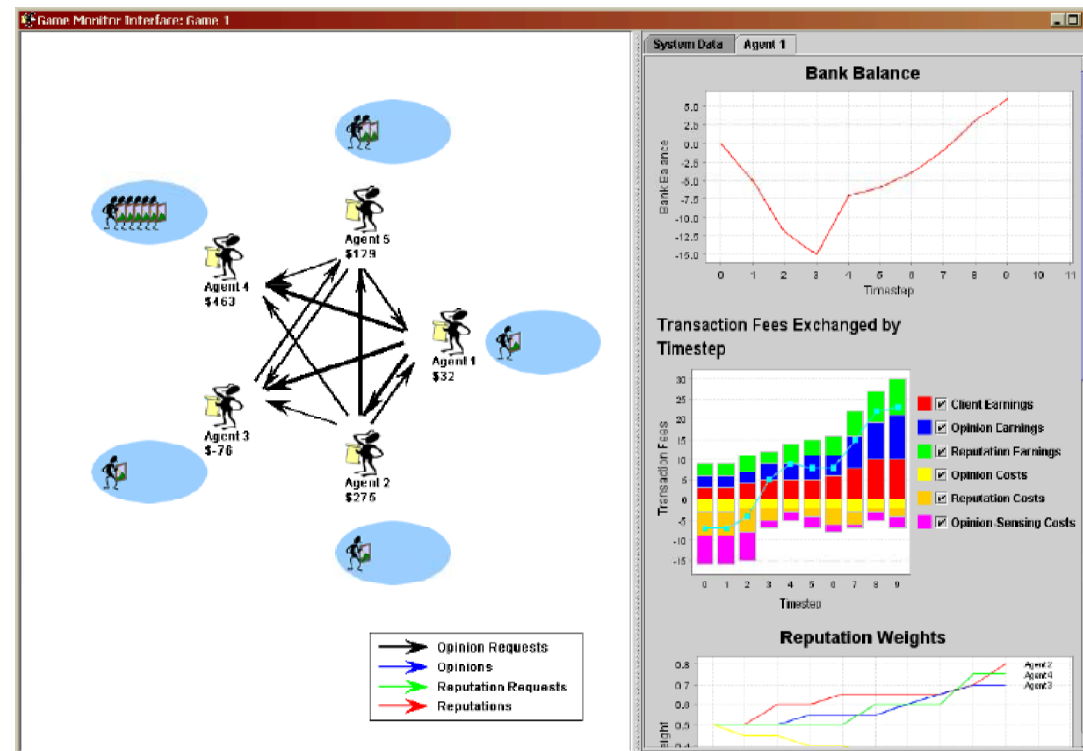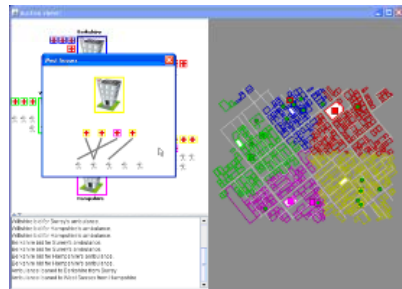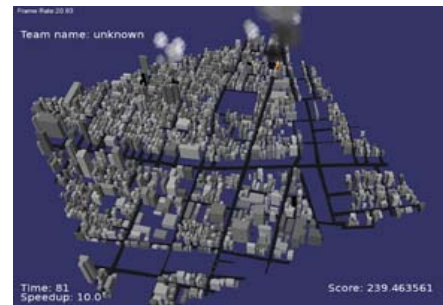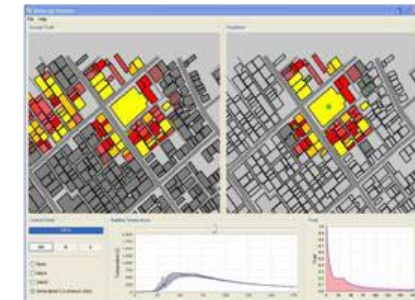
Simulating Trust

- ## RoboCup Rescue Simulation[13]
  - ### An agent simulation system that has been used for trust research[14]
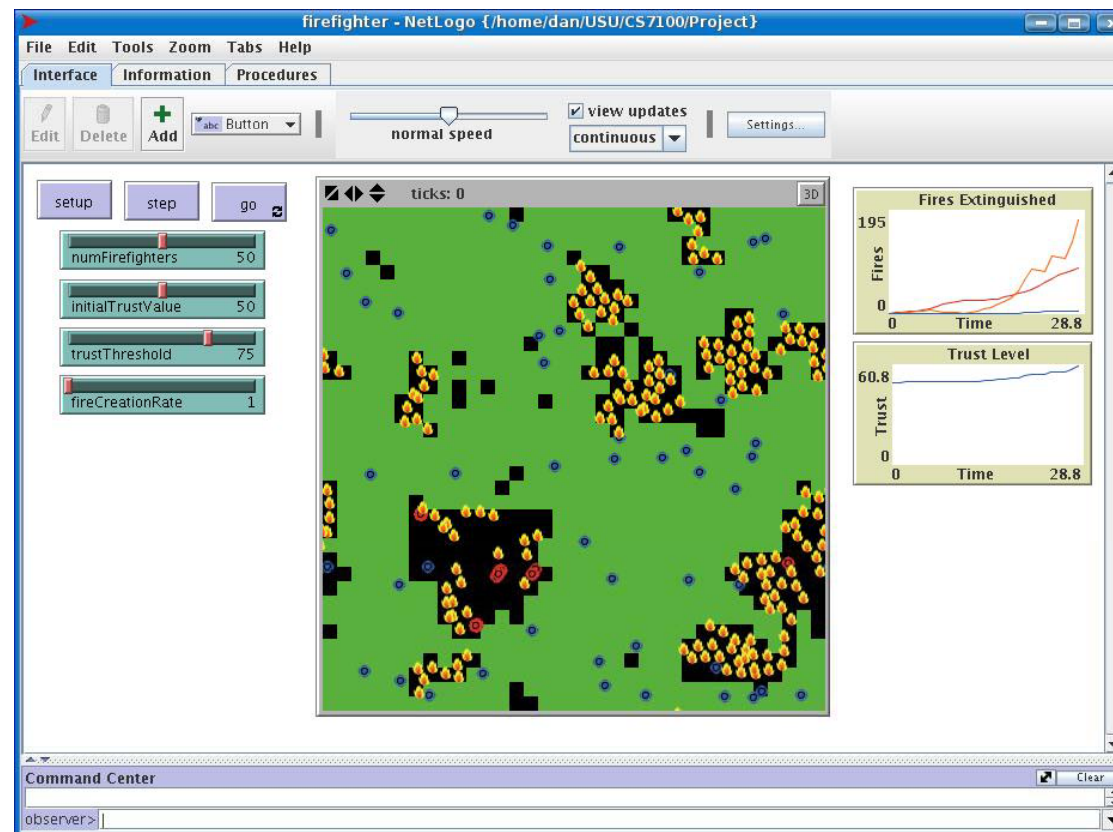

Coalition Formation

Resource Allocation

Estimation with Faulty Sensors

Simulating Trust

- **Firefighting Simulation**
  - **A simulation of human-robot trust relationships**

Human-Robot Trust Game Prototype

- **A prototype of a game for collecting user interaction data was created in NetLogo**
  - **The objective of the prototype was to evaluate playability and design elements for data collection**

Web-based Game

- **To collect data from a variety of users, a web-based game is being developed using the Google Web Toolkit, which provides a variety of tools for collecting data on the backend of a website**

- **The project page, with the latest version of the game and supporting material, can be found at**

https://sites.google.com/a/aggiemail.usu.edu/human-robot-trust-project/
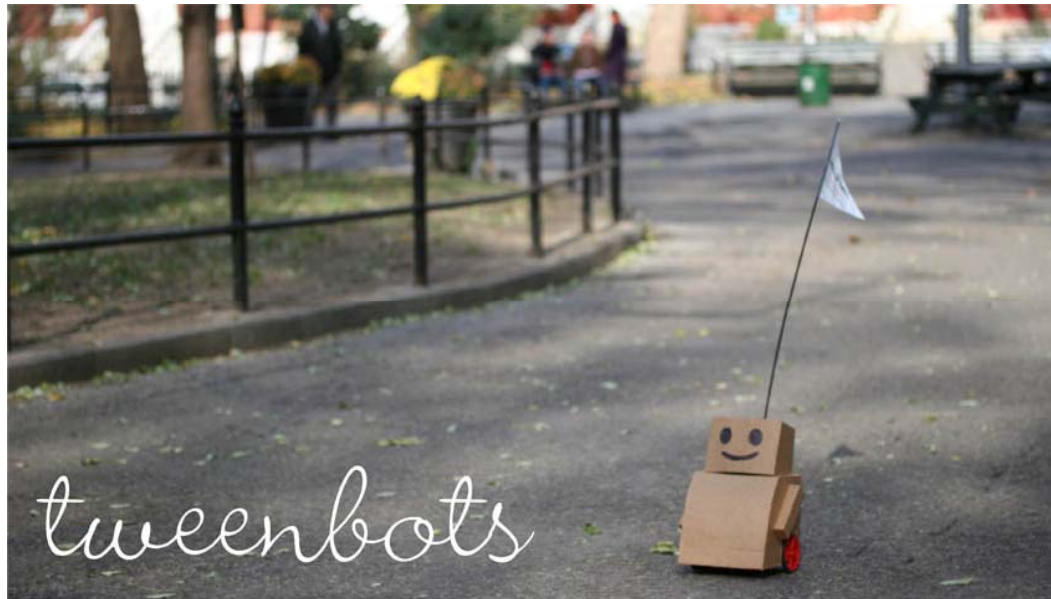
Identifying Human-Agent Trust KPPs

- **The future direction for this project is to develop Key Performance Parameters that can be used by autonomous agent designers**
    - **Serve as design guidelines**
    - **Provide parameters for testing against during development**
    - **Provide validation parameters for field testing**

Trust in Reverse?

tweenbots

A project by Kacie Kinzer:
http://www.tweenbots.com

Trust in Reverse?

Systems & Software Technology Conference 2009

Summary

- **Why Trust Matters**
- **Autonomy Defined**
- **Trust Defined**
- **Research Objective**
- **Identifying Trust Factors**
- **Simulating Trust**
- **Human-Robot Trust Game Prototype**
- **Web-based Game**
- **Identifying Human-Agent Trust KPPs**
- **Trust in Reverse?**

Questions?

OGDEN AIR LOGISTICS CENTER

References

OGDEN AIR LOGISTICS CENTER

1. DARPA, *DARPA Urban Challenge.* Online: [http://www.darpa.mil/GRANDCHALLENGE/]
2. Robert Gates, Secretary Gates Remarks at Maxwell-Gunter Air Force Base, Montgomery Alabama, *DefenseLink News Transcript*, April 21, 2008. Online: [http://www.defenselink.mil/utility/printitem.aspx?print=http://www.defenselink.mil/transcripts/transcript.aspx?transcriptid=4214]
3. OSD, *Unmanned Systems Roadmap, 2007-2032,* December 10, 2007. Online: [http://www.acq.osd.mil/usd/Unmanned%20Systems%20Roadmap.2007-2032.pdf]
4. P.W. Singer, "In the Loop? Armed Robots and the Future of War," *Defense Industry Daily*, January 28, 2009. Online:[http://www.defenseindustrydaily.com/In-the-Loop-Armed-Robots-and-the-Future-of-War-05267/]
5. "Autonomy," *American Heritage Dictionary*, Online: [http://www.answers.com/topic/autonomy]
6. "Autonomous Agent," *Wikipedia*, Online: [http://en.wikipedia.org/wiki/Autonomous_agent]
7. "Autonomous Robot," *Wikipedia*, Online: [http://en.wikipedia.org/wiki/Autonomous_robot]
8. Diego Gambetta, ed., *Trust: Making and Breaking Cooperative Relations,* New York, NY: Basil Blackwell Ltd., 1988. Online: [http://www.nuffield.ox.ac.uk/users/gambetta/gambetta_trust%20book.pdf]
9. Cristiano Castelfranchi and Rino Falcone, "Principles of Trust for MAS: Cognitive Autonomy, Social Importance, and Quantification," Proceedings of ICMAS 1999. Online: [http://aois.org/99/castelfranchi-ICMAS-paper.rtf]
10. Sarvapali Ramchurn, "Devising a Trust Model for Multi-Agent Interactions using Confidence and Reputation," *Applied Artificial Intelligence*, 18:833-852, 2004. Online: [http://users.ecs.soton.ac.uk/nrj/download-files/jaai04.pdf]
11. Karen Fullam and K.Suzanne Barber, "Learning Trust Strategies in Reputation Exchange Networks," *The Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2006),* Hakodate, Japan, May 8-12, 2006, pp. 1241-1248. Online: [http://megatron.iiia.csic.es/art-testbed/pdf/FullamAAMAS2006.pdf]

References

12. *ART Testbed.* Online: [http://www.art-testbed.net/]
13. RoboCup Rescue, *Rescue Agents.* Online: [http://www.robocuprescue.org/agentsim.html]
14. ALADDIN, *Technologies.* Online: [http://www.aladdinproject.org/technologies.html]
15. *tweenbots.* Online: [http://www.tweenbots.com]

OGDEN AIR LOGISTICS CENTER

Acronym List

- **ALCM**   Air Launched Cruise Missile
- **ART**   Agent Reputation and Trust
- **DARPA**   Defense Advanced Research Projects Agency
- **EOD**   Explosive Ordnance Disposal
- **ISS**   International Space Station
- **KPP**   Key Performance Parameter
- **LOAC**   Law Of Armed Conflict
- **NIST**   National Institute for Standards and Technology
- **OSD**   Office of the Secretary of Defense
- **SMXS**   Software Maintenance Squadron
- **UAS**   Unmanned Aerial System
- **UAV**   Unmanned Aerial Vehicle
- **UGV**   Unmanned Ground Vehicle
- **USV**   Unmanned Submersible/Surface Vehicle